Vector-Valued Support Vector Regression (Paper #1751)

Mark James Brudnak, Ph.D.
U.S. Army Tank Automotive Research
Development and Engineering Center

15699

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **18 JUL 2006**	2. REPORT TYPE **N/A**	3. DATES COVERED
4. TITLE AND SUBTITLE **Vector-Valued Support Vector Regression**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Brudnak /Mark James, Ph.D.**		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **USATACOM 6501 E 11 Mile Road Warren, MI 48397-5000**		8. PERFORMING ORGANIZATION REPORT NUMBER **15999**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **TACOM TARDEC**		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited.**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**34**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Vector-Valued Support Vector Regression

- The training data are of the form $\{(\mathbf{x}_i, \mathbf{y}_i)\}_\ell$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$
- $\hat{\mathbf{y}} = \mathbf{f}(\mathbf{x},\pi)$, $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$
- Desire to minimize the loss functional (also called empirical risk functional)

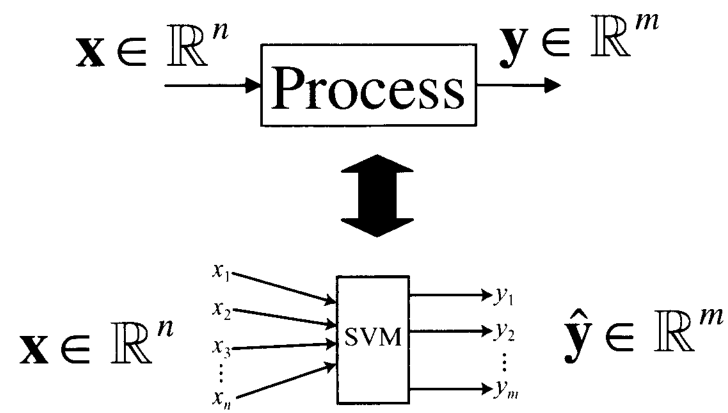
$$\mathcal{J}(\boldsymbol{\pi}) = \sum_{i=1}^{\ell} L\left(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi})\right)$$

- Parameter preference expressed by regularization functional

$$\mathcal{P}(\boldsymbol{\pi})$$

- Balance the two as regulaized risk functional

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C\mathcal{J}(\boldsymbol{\pi})$$



$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C\sum_{i=1}^{\ell} L\left(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi})\right)$$

How should these be defined?

Parameters and estimator

Scalar Case: $y \in \mathbb{R}$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + \sum_{i=1}^{\ell} L\left(y_i, \text{[redacted]}(\mathbf{x}_i, \boldsymbol{\pi})\right)$$

Parameters:

$$\boldsymbol{\pi} = \{\mathbf{w}, b\}, \quad \mathbf{w} \in \mathbb{R}^{\nu}, b \in \mathbb{R}$$

- The weight $\mathbf{w}$ is <u>free</u>.
- The bias b is <u>free</u>.

Estimator form:

[redacted]

- The mapping $\varphi(\cdot)$ is <u>given</u>.
- The estimator is <u>linear</u> in the range of $\varphi(\cdot)$!

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + \sum_{i=1}^{\ell} L\left(\mathbf{y}_i, \text{[redacted]}(\mathbf{x}_i, \boldsymbol{\pi})\right)$$

Parameters:

$$\boldsymbol{\pi} = \{\mathbf{W}, \mathbf{b}\}, \quad \mathbf{W} \in \mathbb{R}^{m \times \nu}, \mathbf{b} \in \mathbb{R}^m$$

- The weight $\mathbf{W}$ is <u>free</u>.
- The bias $\mathbf{b}$ is <u>free</u>.

Estimator form:

[redacted]

- The mapping $\varphi(\cdot)$ is <u>given</u>.
- The estimator is <u>linear</u> in the range of $\varphi(\cdot)$!

Regularization and Loss

Scalar Case: $y \in \mathbb{R}$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} \text{[redacted]}\left(y_i, f(\mathbf{x}_i, \boldsymbol{\pi}) \right)$$

Regularization functional:

$$\mathcal{P}(\boldsymbol{\pi}) = \tfrac{1}{2} \mathbf{w}^T \mathbf{w}$$

- The weight $\mathbf{w}$ penalized.
- The bias b is not penalized.

Loss Function:

[redacted]

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} \text{[redacted]}\left(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi}) \right)$$

Regularization functional:

$$\mathcal{P}(\boldsymbol{\pi}) = \tfrac{1}{2} \mathrm{Tr}\left(\mathbf{W} \mathbf{W}^T \right)$$

- The weight $\mathbf{W}$ penalized.
- The bias $\mathbf{b}$ is not penalized.

Loss Function:

[redacted]

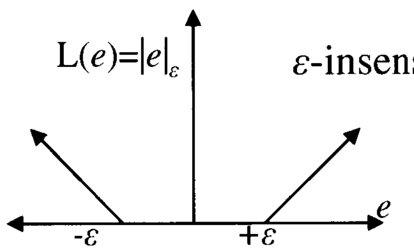
More to follow

The Loss Function (continued)

Scalar Case: $y \in \mathbb{R}$

$$e \triangleq y - \hat{y}$$

$$L(e) \triangleq |e|_{\varepsilon} = \max\left(0, |e| - \varepsilon\right)$$



ε-insensitive Loss

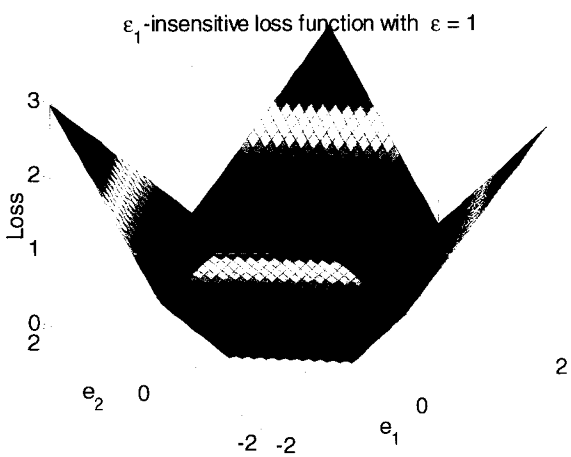
Vector Case: $\mathbf{y}_i \in \mathbb{R}^m$

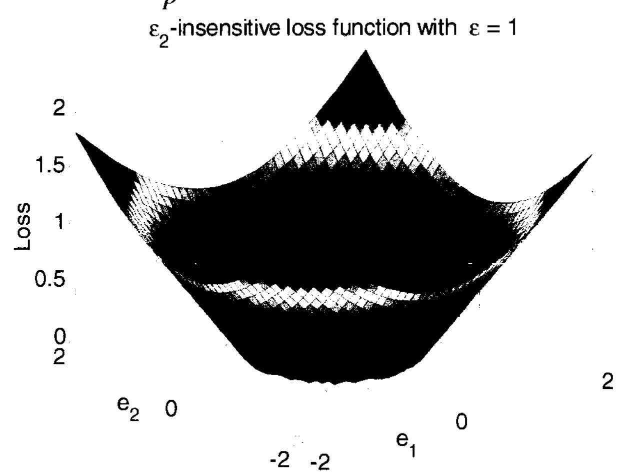
$$\mathbf{e} \triangleq \mathbf{y} - \hat{\mathbf{y}}$$

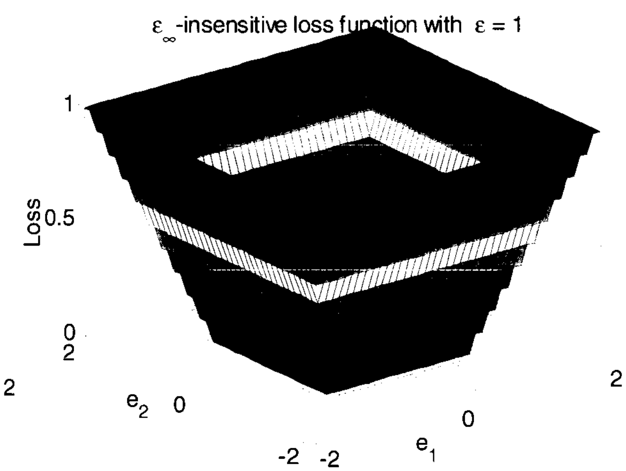
$$L(\mathbf{e}) \triangleq \left| \|\mathbf{e}\|_p \right|_{\varepsilon} = \max\left(0, \|\mathbf{e}\|_p - \varepsilon\right)$$

$$\|\mathbf{e}\|_p \triangleq \begin{cases} \left(\sum_{j=1}^{m} |e_j|^p\right)^{\frac{1}{p}} & 1 \le p < \infty \\ \max\limits_{j}\left(|e_j|\right) & p \sim \infty \end{cases}$$

ε_p-insensitive Loss







Put it all Together

Scalar Case: $y \in \mathbb{R}$

Regularized Risk Functional:

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} L\left(y_i, f\left(\mathbf{x}_i, \boldsymbol{\pi}\right)\right)$$

$$\Downarrow$$

$$R_{reg} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\left|y_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) - b\right|_{\varepsilon}$$

Problem:

Minimize: $\underset{\{\mathbf{w}, b\}}{\text{Minimize}}\; R_{reg}\left(\mathbf{w}, b\right)$

Given: $\left\{\mathbf{x}_i, y_i\right\}_{i=1}^{\ell}, C, \varepsilon, \boldsymbol{\varphi}(\cdot)$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Regularized Risk Functional:

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} L\left(\mathbf{y}_i, \mathbf{f}\left(\mathbf{x}_i, \boldsymbol{\pi}\right)\right)$$

$$\Downarrow$$

$$R_{reg} = \frac{1}{2}\mathrm{Tr}(\mathbf{W}\mathbf{W}^T) + C\sum_{i=1}^{\ell}\left|\left\|\mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b}\right\|_p\right|_{\varepsilon}$$

Problem:

Minimize: $\underset{\{\mathbf{W}, \mathbf{b}\}}{\text{Minimize}}\; R_{reg}\left(\mathbf{W}, \mathbf{b}\right)$

Given: $\left\{\mathbf{x}_i, \mathbf{y}_i\right\}_{i=1}^{\ell}, C, \varepsilon, \left\|\cdot\right\|_p, \boldsymbol{\varphi}(\cdot)$

Implicitly given

But what about the mapping $\varphi(\cdot)$?

Back to SVM: How to solve it

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

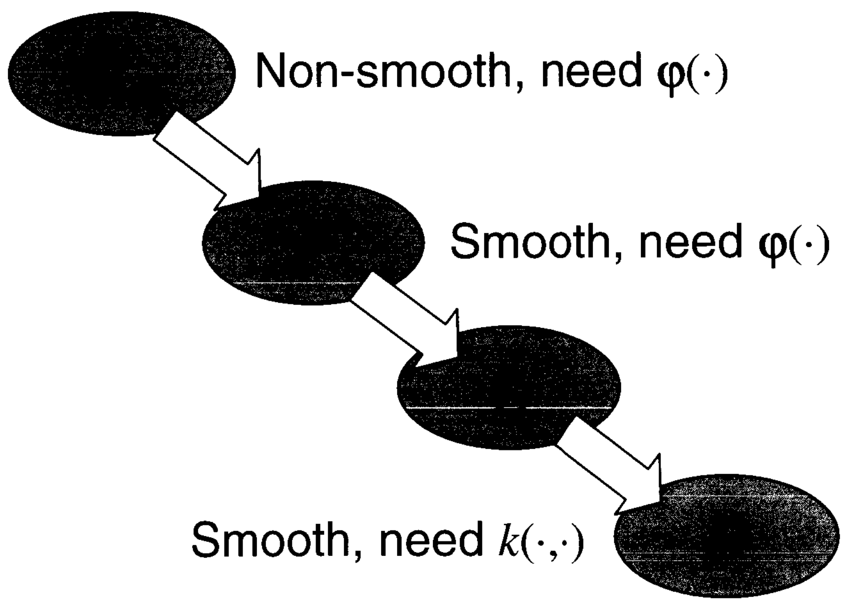
Scalar Case: $y \in \mathbb{R}$

Problem:

Minimize$_{\{\mathbf{w},b\}}$:

$$R_{reg} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\left|y_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) - b\right|_{\varepsilon}$$

Approach:



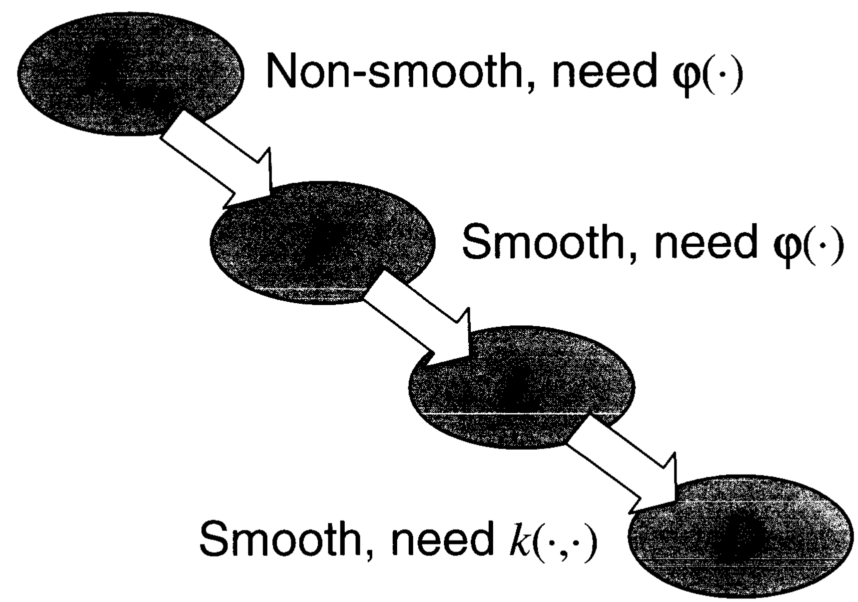
Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Problem:

Minimize$_{\{\mathbf{W},\mathbf{b}\}}$:

$$R_{reg} = \frac{1}{2}\mathrm{Tr}(\mathbf{W}\mathbf{W}^T) + C\sum_{i=1}^{\ell}\left|\left\|\mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b}\right\|_p\right|_{\varepsilon}$$

Approach:



The Primal Problem

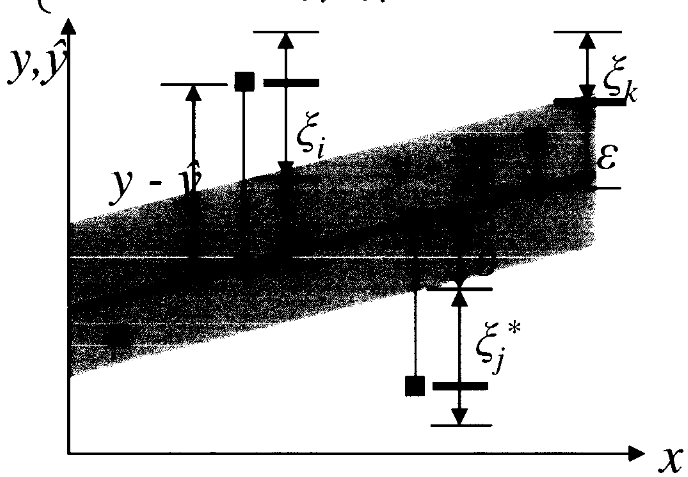
Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Minimize: $\mathbf{w}, b, \{\xi_i, \xi_i^*\}_{i=1}^{\ell}$

$$P = \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*)$$

subject to:

$$\begin{cases} y_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) - b - \varepsilon - \xi_i \le 0 \\ -y_i + \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) + b - \varepsilon - \xi_i^* \le 0 \\ \xi_i, \xi_i^* > 0 \end{cases}$$



Minimize: $\mathbf{W}, \mathbf{b}, \{\xi_i, \boldsymbol{\delta}_i, \boldsymbol{\delta}_i^*\}_{i=1}^{\ell}$

Primal Variables

$$P = \tfrac{1}{2}\mathrm{Tr}\left(\mathbf{W}\mathbf{W}^T\right) + C\sum_{i=1}^{\ell}\xi_i$$

subject to:

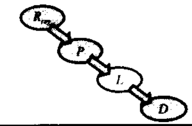
$$\begin{cases} \left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p - \varepsilon - \xi_i \le 0, \quad \xi_i \ge 0 \\ \mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i \le \mathbf{0}, \quad \boldsymbol{\delta}_i \ge \mathbf{0} \\ -\mathbf{y}_i + \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^* \le \mathbf{0}, \quad \boldsymbol{\delta}_i^* \ge \mathbf{0} \end{cases}$$

$\boldsymbol{\delta}_i$, $\boldsymbol{\delta}_i^*$ and ξ_i are slack variables

Example: slack variable π

$$a \le b \Leftrightarrow \begin{cases} a + \pi = b \\ \pi \ge 0 \end{cases}$$

The Lagrange Problem



Scalar Case: $y \in \mathbb{R}$

Minimize, $\mathbf{w}, b, \{\xi_i, \xi_i^*\}_{i=1}^{\ell}$ Maximize: $\{\alpha_i, \alpha_i^*, \eta_i, \eta_i^*\}_{i=1}^{\ell}$

$$L \triangleq \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*)$$

$$-\sum_{i=1}^{\ell}\alpha_i\left(\varepsilon + \xi_i - y_i + \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) + b\right)$$

$$-\sum_{i=1}^{\ell}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) - b)$$

$$-\sum_{i=1}^{\ell}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$

subject to: $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Minimize, $\mathbf{W}, \mathbf{b}, \{\xi_i, \boldsymbol{\delta}_i, \boldsymbol{\delta}_i^*\}_{i=1}^{\ell}$ Maximize: $\{\alpha_i, \eta_i, \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i^*, \boldsymbol{\theta}_i, \boldsymbol{\theta}_i^*\}_{i=1}^{\ell}$

Dual Variables

$$L \triangleq \tfrac{1}{2}\mathrm{Tr}\left(\mathbf{W}\mathbf{W}^T\right) + C\sum_{i=1}^{\ell}\xi_i$$

$$+\sum_{i=1}^{\ell}\alpha_i\left(\left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p - \varepsilon - \xi_i\right) - \sum_{i=1}^{\ell}\eta_i\xi_i$$

$$-\sum_{i=1}^{\ell}\boldsymbol{\gamma}_i^T\left(\mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i\right)$$

$$-\sum_{i=1}^{\ell}\boldsymbol{\gamma}_i^{*T}\left(-\mathbf{y}_i + \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^*\right)$$

$$-\sum_{i=1}^{\ell}(\boldsymbol{\theta}_i^T\boldsymbol{\delta}_i + \boldsymbol{\theta}_i^{*T}\boldsymbol{\delta}_i^*)$$

subject to: $\alpha_i, \eta_i \geq 0, \quad \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i^*, \boldsymbol{\theta}_i, \boldsymbol{\theta}_i^* \geq \mathbf{0}$

Minimization of L over Primals

Scalar Case: $y \in \mathbb{R}$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{\ell} (\alpha_i^* - \alpha_i) = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{\ell} (\alpha_i^* - \alpha_i)\mathbf{x}_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0$$

Introduce: $\beta_i = \alpha_i - \alpha_i^*$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\frac{\partial L}{\partial \mathbf{b}} = \sum_{i=1}^{\ell} \left(\boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_i\right) = 0$$

$$\frac{\partial L}{\partial \mathbf{W}} = \mathbf{W} - \sum_{i=1}^{\ell} \left(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\right) \boldsymbol{\varphi}^T(\mathbf{x}_i) = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0$$

$$\frac{\partial L}{\partial \boldsymbol{\delta}_i} = \alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\| \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \right\|_p} \right)^{p-1} - \boldsymbol{\gamma}_i - \boldsymbol{\theta}_i = 0$$

$$\frac{\partial L}{\partial \boldsymbol{\delta}_i^*} = \alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\| \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \right\|_p} \right)^{p-1} - \boldsymbol{\gamma}_i^* - \boldsymbol{\theta}_i^* = 0$$

Introduce: $\boldsymbol{\Gamma}_i = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*$

$\mathbf{e}_i = \boldsymbol{\delta}_i - \boldsymbol{\delta}_i^*$

Implications of Optimization

Scalar Case: $y \in \mathbb{R}$

Equality Constraint:

$$\sum_{i=1}^{\ell} \beta_i = 0$$

Weights:

[redacted] $= \sum_{i=1}^{\ell} \beta_i \boldsymbol{\varphi}(\mathbf{x}_i)$

Constraints:

$$C - \alpha_i - \eta_i = 0 \quad \rightarrow \quad \alpha_i \leq C$$

$$C - \alpha_i^* - \eta_i^* = 0 \quad \rightarrow \quad \alpha_i^* \leq C$$

Estimator:

$\hat{y} =$ [redacted] $^T \boldsymbol{\varphi}(\mathbf{x}) + b =$ [redacted] $+ b$

$= \sum_{i=1}^{\ell} \beta_i$ [redacted] $+ b$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Equality Constraint:

$$\sum_{i=1}^{\ell} \boldsymbol{\Gamma}_i = 0$$

Weights:

[redacted] $= \sum_{i=1}^{\ell} \boldsymbol{\Gamma}_i \boldsymbol{\varphi}^T(\mathbf{x}_i)$

Constraints:

$$C - \alpha_i - \eta_i = 0 \quad \rightarrow \quad \alpha_i \leq C$$

Estimator:

$\hat{\mathbf{y}}(\mathbf{x}) =$ [redacted] $\boldsymbol{\varphi}(\mathbf{x}) + \mathbf{b} =$ [redacted] $+ \mathbf{b}$

$= \sum_{i=1}^{\ell} \boldsymbol{\Gamma}_i$ [redacted] $+ \mathbf{b}$

The Dual Problem

Scalar Case: $y \in \mathbb{R}$

Regularized Risk:

Minimize: $\{\mathbf{w}, b\}$

$$R_{reg} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\left|y_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) - b\right|_{\varepsilon}$$

Dual Problem:

Maximize: $\{\beta_i\}_{i=1}^{\ell}$

$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\beta_i\beta_j$ [redacted] $+ \sum_{i=1}^{\ell} y_i\beta_i - \varepsilon\sum_{i=1}^{\ell}|\beta_i|$

Subject to:

$$\sum_{i=1}^{\ell}\beta_i = 0, \quad |\beta_i| \le C$$

Solve for β_i numerically

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Regularized Risk:

Minimize: $\{\mathbf{W}, \mathbf{b}\}$

$$R_{reg} = \frac{1}{2}\mathrm{Tr}(\mathbf{W}\mathbf{W}^T) + C\sum_{i=1}^{\ell}\left|\left\|\mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b}\right\|_p\right|_{\varepsilon}$$

Dual Problem:

Maximize: $\{\boldsymbol{\Gamma}_i\}_{i=1}^{\ell}$

$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\boldsymbol{\Gamma}_i^T\boldsymbol{\Gamma}_j$ [redacted] $+ \sum_{i=1}^{\ell}\mathbf{y}_i^T\boldsymbol{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\|\boldsymbol{\Gamma}_i\|_q$

Subject to:

$$\sum_{i=1}^{\ell}\boldsymbol{\Gamma}_i = 0, \quad \|\boldsymbol{\Gamma}_i\|_q \le C \qquad \frac{1}{p} + \frac{1}{q} = 1$$

Solve for Γ_i numerically

Half Way There

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Dual Problem: Solve for β_i numerically

Maximize: $D\left(\{\beta_i\}_{i=1}^{\ell}\right)$

Subject to: $\sum_{i=1}^{\ell} \beta_i = 0, \quad |\beta_i| \le C$

Estimator:

$$\hat{y} = \sum_{i=1}^{\ell} \beta_i k\left(\mathbf{x}_i, \mathbf{x}\right) + b$$

Support Vectors: $\{\mathbf{x}_i : \beta_i \neq 0\}$

The bias remains to be found

Dual Problem: Solve for Γ_i numerically

Maximize: $D\left(\{\mathbf{\Gamma}_i\}_{i=1}^{\ell}\right)$

Subject to: $\sum_{i=1}^{\ell} \mathbf{\Gamma}_i = 0, \quad \|\mathbf{\Gamma}_i\|_q \le C$

Estimator:

$$\hat{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{\ell} \mathbf{\Gamma}_i k\left(\mathbf{x}_i, \mathbf{x}\right) + \mathbf{b}$$

Support Vectors: $\{\mathbf{x}_i : \mathbf{\Gamma}_i \neq \mathbf{0}\}$

The bias remains to be found

Must Develop the KKT Conditions to find bias

KKT Conditions

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$
$\alpha_i(\varepsilon + \xi_i - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) = 0$	$\alpha_i\left(\left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p - \varepsilon - \xi_i\right) = 0$
$\alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b) = 0$	$\eta_i \xi_i = (C - \alpha_i)\xi_i = 0$
$\eta_i \xi_i = (C - \alpha_i)\xi_i = 0$	$\boldsymbol{\gamma}_i \circ \left(\mathbf{y}_i - \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i\right) = \mathbf{0}$
$\eta_i^* \xi_i^* = (C - \alpha_i^*)\xi_i^* = 0$	$\boldsymbol{\gamma}_i^* \circ \left(-\mathbf{y}_i + \mathbf{W}\boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^*\right) = \mathbf{0}$
	$\boldsymbol{\theta}_i \circ \boldsymbol{\delta}_i = \left(\alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p}\right)^{p-1} - \boldsymbol{\gamma}_i^*\right) \circ \boldsymbol{\delta}_i = \mathbf{0}$
Note: $\circ$ denotes a parallel or element-wise product. Exponents are taken element-wise	$\boldsymbol{\theta}_i^* \circ \boldsymbol{\delta}_i^* = \left(\alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p}\right)^{p-1} - \boldsymbol{\gamma}_i^*\right) \circ \boldsymbol{\delta}_i^* = \mathbf{0}$

Implications of KKT Conditions

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Primal + Dual + KKT conditions imply

(absolute value and exponent are taken element wise)

$$\boxed{\boldsymbol{\delta}_i \circ \boldsymbol{\delta}_i^* = \mathbf{0}} \quad \boxed{\boldsymbol{\gamma}_i \circ \boldsymbol{\gamma}_i^* = \mathbf{0}} \quad \boxed{\Gamma_{i,j} \neq 0 \quad \rightarrow \quad \operatorname{sign}\left(\Gamma_{i,j}\right) = \operatorname{sign}\left(e_{i,j}\right)}$$

$$\boxed{\left|\boldsymbol{\Gamma}_i\right| = \alpha_i \frac{d}{d\mathbf{e}_i}\left(\left\|\mathbf{e}_i\right\|_p\right)}$$

[redacted]

Lemma 1

$$\boxed{\alpha_i = \left\|\boldsymbol{\Gamma}_i\right\|_q} \quad \boxed{\boldsymbol{\delta}_i - \boldsymbol{\delta}_i^* \equiv \mathbf{e}_i}$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$\boxed{\frac{\left|\boldsymbol{\Gamma}_i\right|}{\left\|\boldsymbol{\Gamma}_i\right\|_q} = \left|\frac{\mathbf{e}_i}{\left\|\mathbf{e}_i\right\|_p}\right|^{p-1}, \quad \forall p \in [1, \infty]}$$

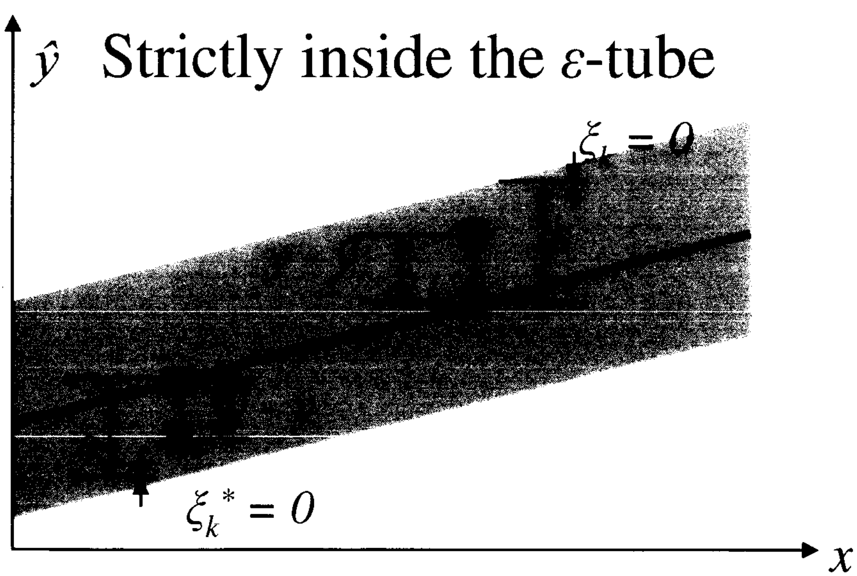
or

$$\boxed{\left|\frac{\boldsymbol{\Gamma}_i}{\left\|\boldsymbol{\Gamma}_i\right\|_q}\right|^q = \left|\frac{\mathbf{e}_i}{\left\|\mathbf{e}_i\right\|_p}\right|^p, \quad p \neq 1}$$

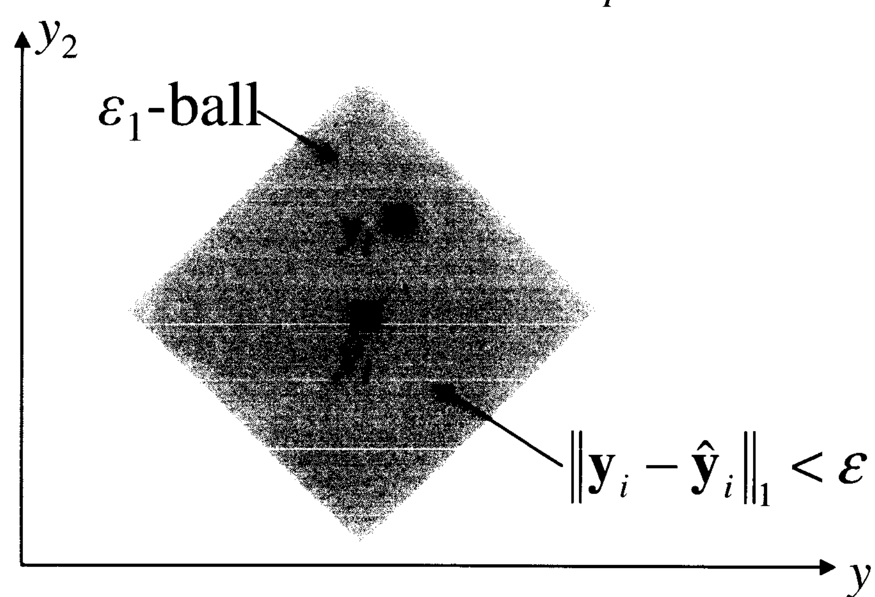
KKT Conditions (in the tube)

$$\frac{1}{p}+\frac{1}{q}=1$$

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$
$\alpha_i = \alpha_i^* = 0$	$\alpha_i = \lVert \boldsymbol{\Gamma}_i \rVert_q = 0$
$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b = e_i < \varepsilon$	$\Rightarrow \lVert \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \rVert_p = \lVert \mathbf{e}_i \rVert_p < \varepsilon$
$\Rightarrow - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b = -e_i < \varepsilon$	



Strictly inside the ε_p-ball



KKT Conditions (on the margin)

$\frac{1}{p}+\frac{1}{q}=1$

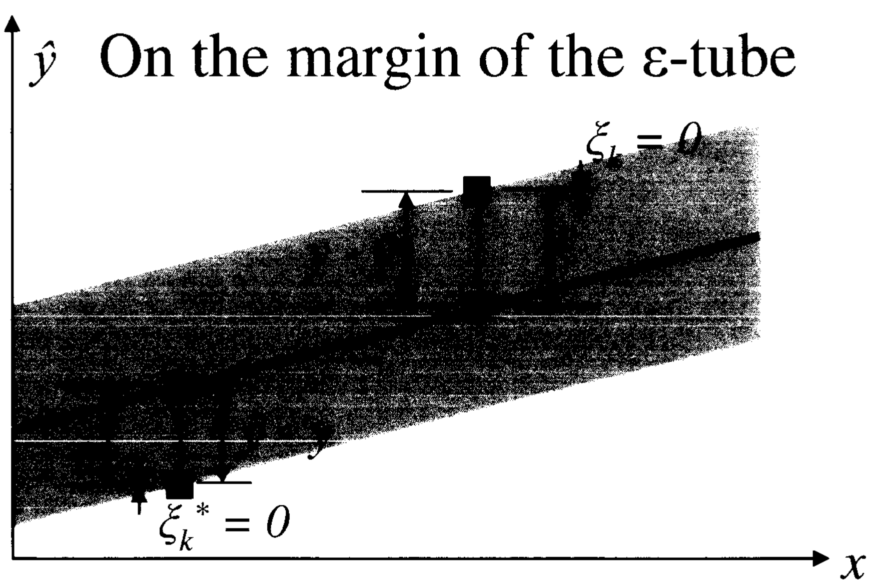
Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$\alpha_i \in (0, C)$

$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b \triangleq e_i = \varepsilon$

[redacted]

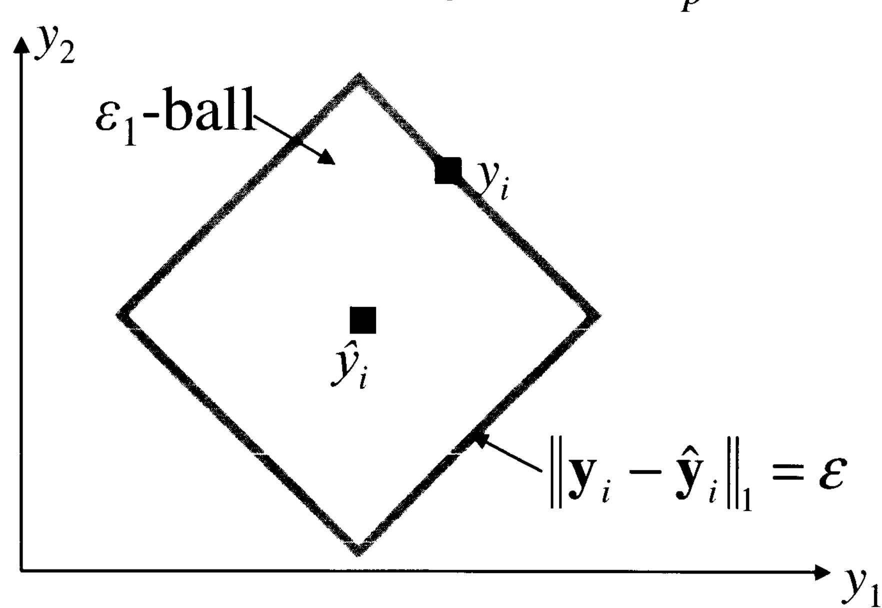
$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b \triangleq$ [redacted]



$\alpha_i = \left\|\boldsymbol{\Gamma}_i\right\|_q \in (0, C)$

$\Rightarrow \left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p = \left\|\mathbf{e}_i\right\|_p = \varepsilon$

On the boundary of the ε_p-ball



KKT Conditions (out of the tube)

$$\frac{1}{p}+\frac{1}{q}=1$$

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

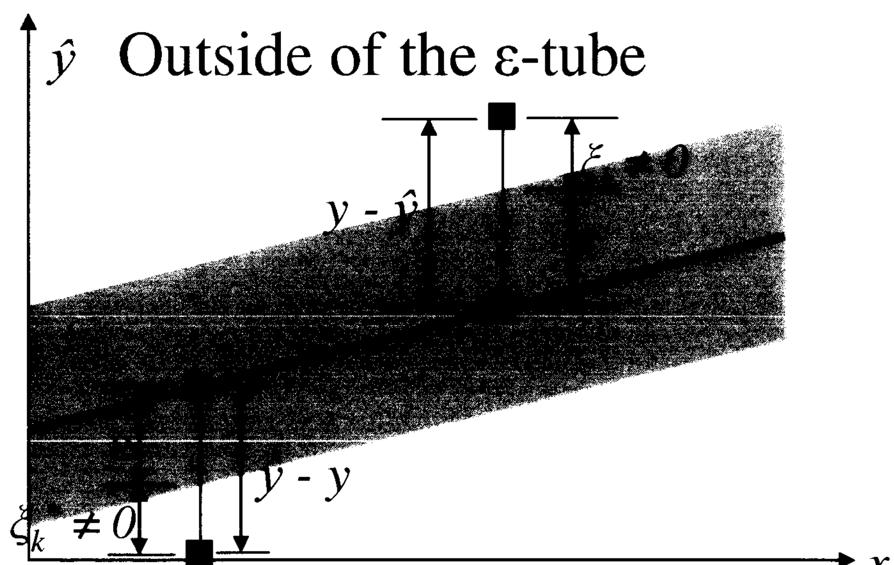
Scalar Case:

$$\alpha_i = C$$

$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b \triangleq e_i > \varepsilon$$

[redacted]

$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b \triangleq$$ [redacted]

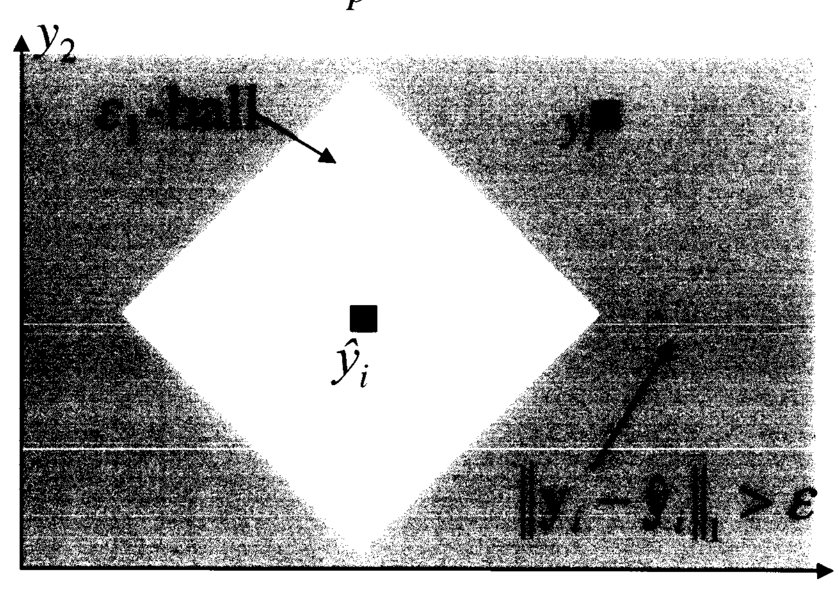


Vector Case:

$$\alpha_i = \left\|\boldsymbol{\Gamma}_i\right\|_q = C$$

$$\Rightarrow \left\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\right\|_p = \left\|\mathbf{e}_i\right\|_p > \varepsilon$$

Out of the ε_p-ball

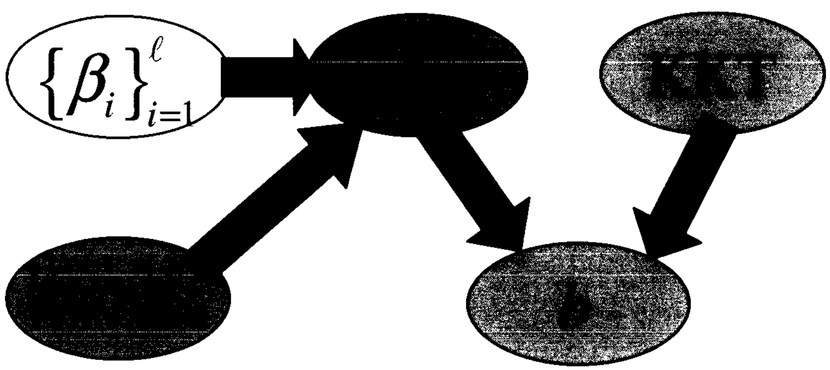


Finding the Bias

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$
Maximize: $D\left(\{\beta_i\}_{i=1}^{\ell}\right)$ Solve for β_i numerically	Maximize: $D\left(\{\mathbf{\Gamma}_i\}_{i=1}^{\ell}\right)$ Solve for Γ_i numerically
Subject to: $\sum_{i=1}^{\ell} \beta_i = 0, \quad \lvert\beta_i\rvert \le C$	Subject to: $\sum_{i=1}^{\ell} \mathbf{\Gamma}_i = 0, \quad \lVert\mathbf{\Gamma}_i\rVert_q \le C$

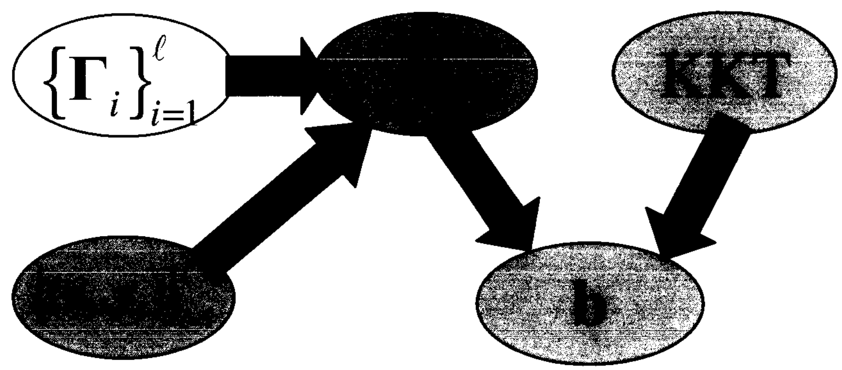
$e_i =$ [redacted] $- \hat{y}_i$

$=$ [redacted] $- \sum_{j=1}^{\ell} \beta_j k($ [redacted] $, \mathbf{x}_i) - b =$ [redacted] $- b$, where the first part (the terms before $-b$) is labelled F_i



$\mathbf{e}_i =$ [redacted] $- \hat{\mathbf{y}}_i$

$=$ [redacted] $- \sum_{j=1}^{\ell} \mathbf{\Gamma}_j k($ [redacted] $, \mathbf{x}_i) - \mathbf{b} =$ [redacted] $- \mathbf{b}$, where the first part (the terms before $-\mathbf{b}$) is labelled $\mathbf{F}_i$



Finding the bias ($1 < p < \infty$)

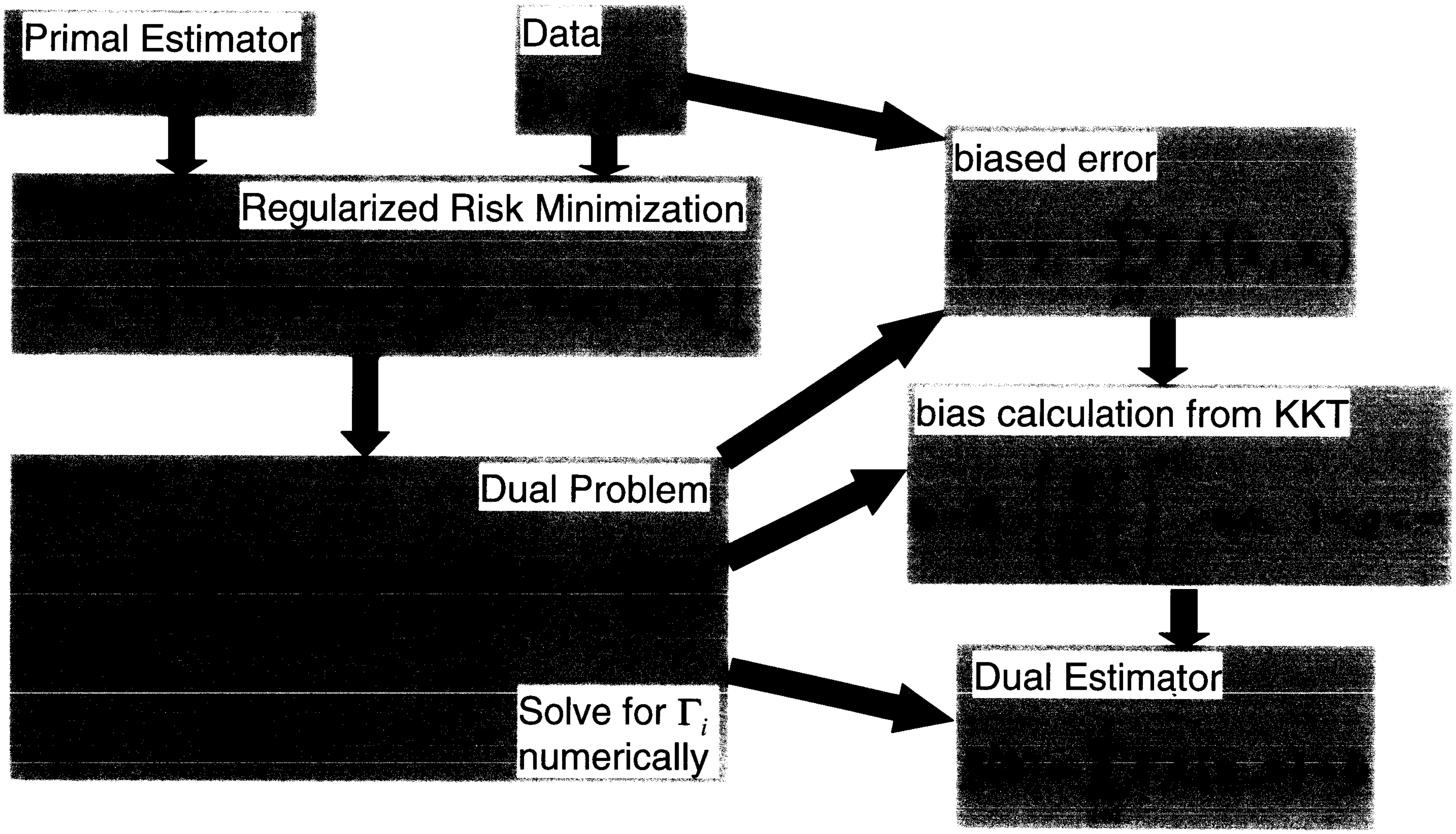
$\frac{1}{p} + \frac{1}{q} = 1$

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$
For $i \in \mathcal{M} = \{i : \lvert\beta_i\rvert \in (0,C)\}$	For $i \in \mathcal{M} \triangleq \{i : \lVert\mathbf{\Gamma}_i\rVert_q \in (0,C)\}$
$F_i - b = e_i = \mathrm{sign}(\beta_i)\varepsilon$ by KKT	$\mathbf{F}_i - \mathbf{b} = \mathbf{e}_i = \left(\frac{\lvert\mathbf{\Gamma}_i\rvert}{\lVert\mathbf{\Gamma}_i\rVert_q}\right)^{q-1} \mathrm{sign}(\mathbf{\Gamma}_i)\varepsilon$ by KKT
if $s_i = \mathrm{sign}(\beta_i)$	if $\boldsymbol{\sigma}_i = \mathrm{sign}(\mathbf{\Gamma}_i)$
it follows that	it follows that
$\boxed{b = F_i - s_i\varepsilon}$	$\boxed{\mathbf{b} = \mathbf{F}_i - \left(\frac{\lvert\mathbf{\Gamma}_i\rvert}{\lVert\mathbf{\Gamma}_i\rVert_q}\right)^{q-1} \circ \boldsymbol{\sigma}_i\varepsilon, \quad 1 < q < \infty}$

Summary

$\frac{1}{p}+\frac{1}{q}=1$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



Comparison

Scalar Case: $y \in \mathbb{R}$	Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Scalar Case

Estimator form:

$$\hat{y}(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) + b = \sum_{i \in \mathcal{SV}} \beta_i k(\mathbf{x}_i, \mathbf{x}) + b$$

Optimization Problem:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \beta_i \beta_j k\left(\mathbf{x}_i, \mathbf{x}_j\right) + \sum_{i=1}^{\ell} y_i \beta_i - \varepsilon \sum_{i=1}^{\ell} |\beta_i|$$

Subject to: $\sum_{i=1}^{\ell} \beta_i = 0, \quad |\beta_i| \le C$

KKT:

$$\begin{aligned}
\beta_i = 0 &\quad\rightarrow\quad |e_i| < \varepsilon \\
|\beta_i| \in (0, C) &\quad\rightarrow\quad |e_i| = \varepsilon \\
|\beta_i| = C &\quad\rightarrow\quad |e_i| > \varepsilon
\end{aligned}$$

Vector Case

Estimator form:

$$\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}) + \mathbf{b} = \sum_{i \in \mathcal{SV}} \boldsymbol{\Gamma}_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}$$

Optimization Problem:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_j k\left(\mathbf{x}_i, \mathbf{x}_j\right) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \boldsymbol{\Gamma}_i - \varepsilon \sum_{i=1}^{\ell} \|\boldsymbol{\Gamma}_i\|_q$$

Subject to: $\sum_{i=1}^{\ell} \boldsymbol{\Gamma}_i = 0, \quad \|\boldsymbol{\Gamma}_i\|_q \le C$

KKT:

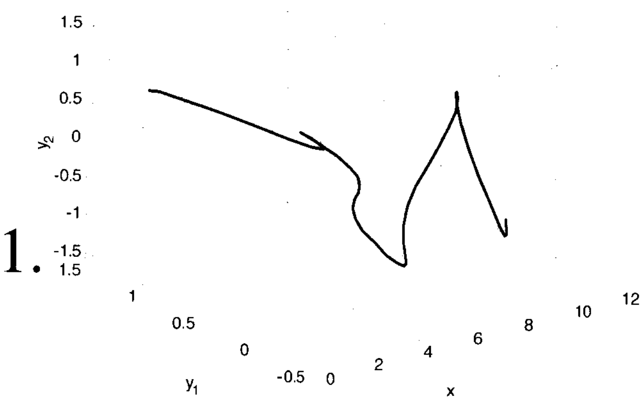
$$\begin{aligned}
\|\boldsymbol{\Gamma}_i\|_q = 0 &\quad\rightarrow\quad \|\mathbf{e}_i\|_p < \varepsilon \\
\|\boldsymbol{\Gamma}_i\|_q \in (0, C) &\quad\rightarrow\quad \|\mathbf{e}_i\|_p = \varepsilon \\
\|\boldsymbol{\Gamma}_i\|_q = C &\quad\rightarrow\quad \|\mathbf{e}_i\|_p > \varepsilon
\end{aligned}$$

Demonstrations for $p = 1, 2, \infty$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

We will demonstrate the VV-SVR for the following process:

$$\mathbf{y}(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} e^{0.1x}\,\mathrm{sinc}(x) \\ \cos(0.1x^2) \end{bmatrix}$$



The following RBF kernel was used with $\sigma = 1$.

$$k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2} \right)$$

50 samples were chosen on the interval [0, 10] for training and $\varepsilon = 0.1$.

Matlab was used as the solver.

The 1-Norm

$p = 1$

$q \sim \infty$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\mathbf{\Gamma}_i^T\mathbf{\Gamma}_j k\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{y}_i^T\mathbf{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\left\|\mathbf{\Gamma}_i\right\|_\infty$$

Subject to:

$$\sum_{i=1}^{\ell}\mathbf{\Gamma}_i = 0, \quad \left\|\mathbf{\Gamma}_i\right\|_\infty \le C$$

Note: D is non-smooth in its objective and constraint.

The dual problem after introduction of α_i:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\mathbf{\Gamma}_i^T\mathbf{\Gamma}_j k\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{y}_i^T\mathbf{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\alpha_i$$

Subject to:

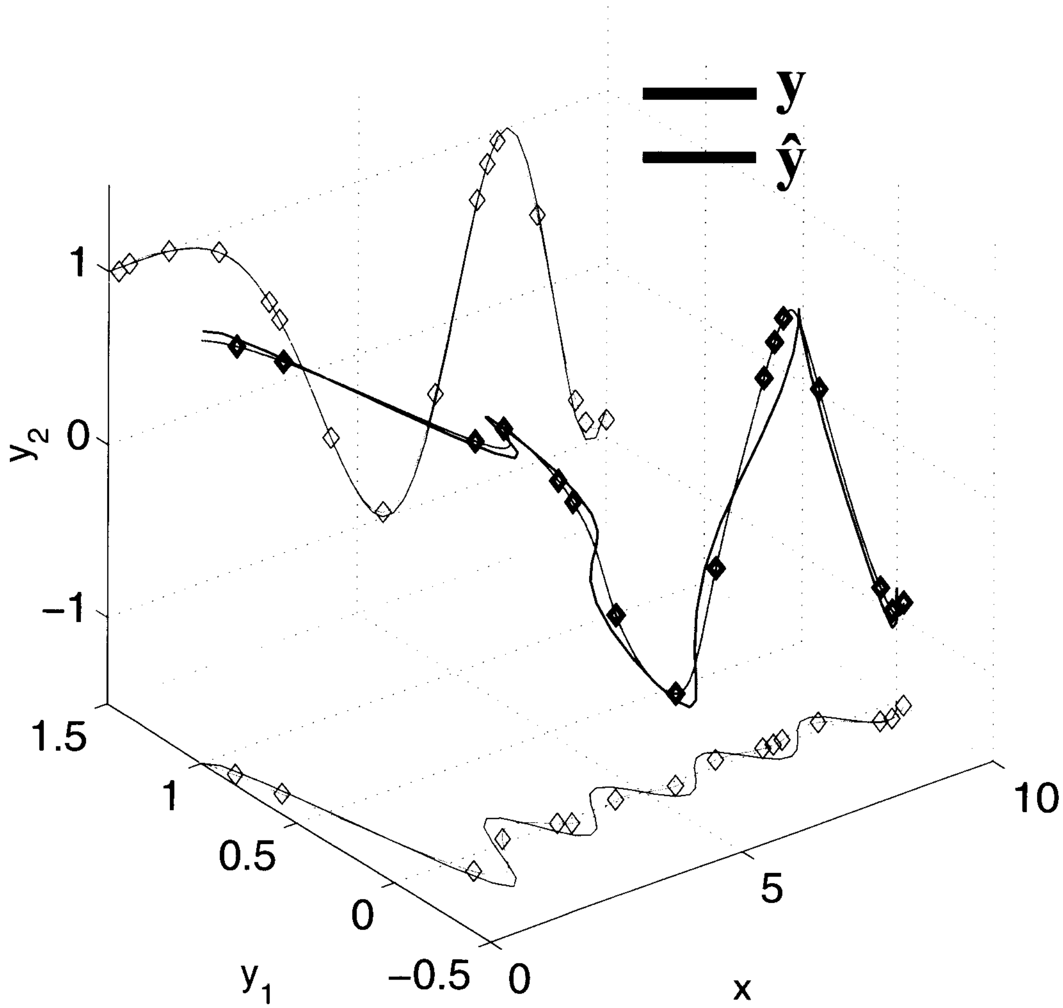
$$\sum_{i=1}^{\ell}\mathbf{\Gamma}_i = 0, \quad -\alpha_i\mathbf{1} \le \mathbf{\Gamma}_i \le \alpha_i\mathbf{1}, \quad \alpha_i \le C$$

Note: D is quadratic in its objective and linear in its constraint.

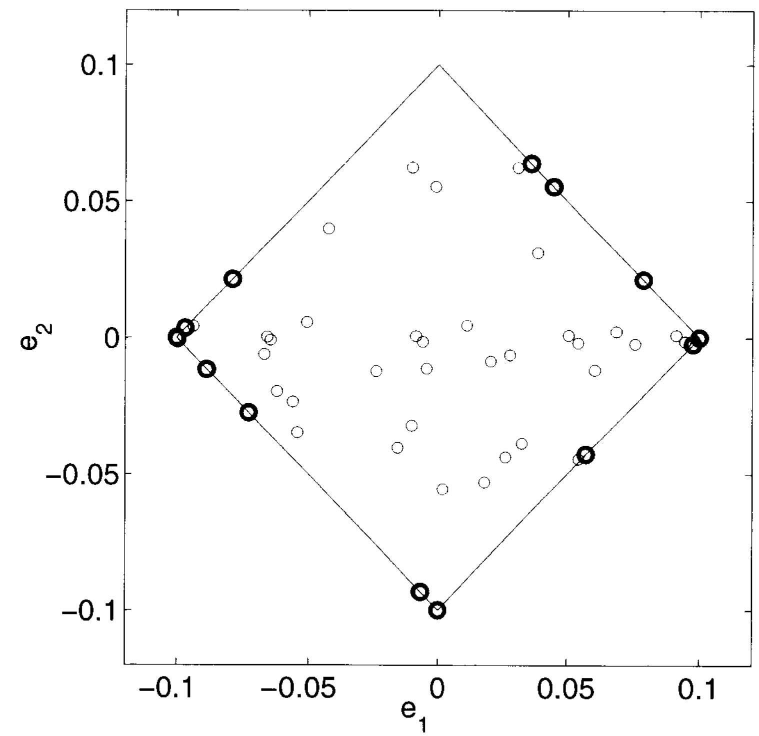
Solve for Γ_i using a standard QP package.

Demonstration (1-Norm)

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



16 support vectors



Matlab's `quadprog()` was used to find Γ_i.

The 2-Norm

$p = 2$
$q = 2$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\mathbf{\Gamma}_i^T\mathbf{\Gamma}_j k\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{y}_i^T\mathbf{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\left\|\mathbf{\Gamma}_i\right\|_2$$

Subject to:

$$\sum_{i=1}^{\ell}\mathbf{\Gamma}_i = 0, \quad \left\|\mathbf{\Gamma}_i\right\|_2 \leq C$$

Note: D is non-smooth in its objective and constraint.
The constraint is nonlinear.

The dual problem after introduction of α_i:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\mathbf{\Gamma}_i^T\mathbf{\Gamma}_j k\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{y}_i^T\mathbf{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\alpha_i$$

Subject to:

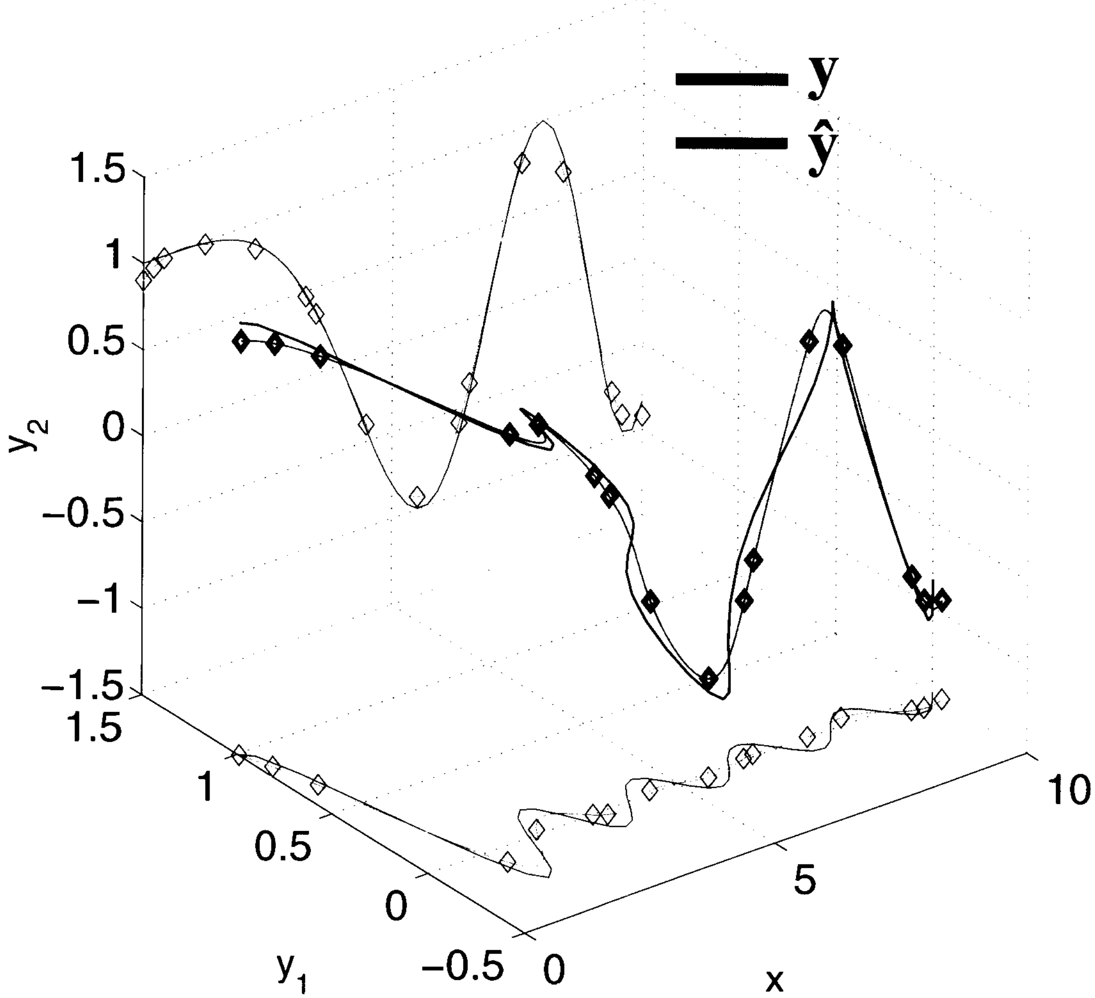
$$\sum_{i=1}^{\ell}\mathbf{\Gamma}_i = 0, \quad \mathbf{\Gamma}_i^T\mathbf{\Gamma}_i = \alpha_i^2, \quad 0 \leq \alpha_i \leq C$$

Note: D is quadratic in its objective and nonlinear but smooth in its constraint.

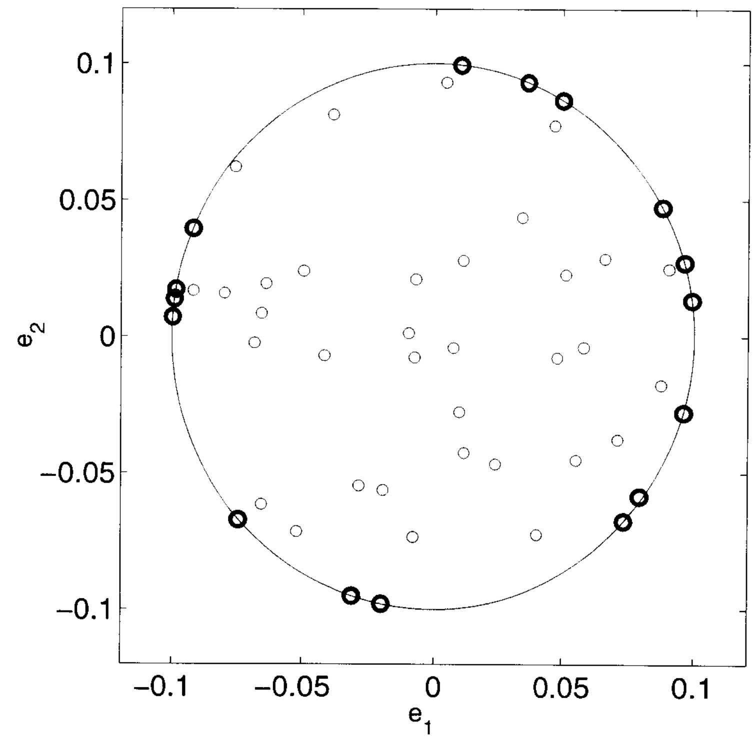
Solve for Γ_i using a standard nonlinear programming package which can use gradients.

Demonstration (2-Norm)

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



16 support vectors



Matlab's `fmincon()` was used to find Γ_i.

The ∞-Norm

$p \sim \infty$

$q = 1$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\boldsymbol{\Gamma}_i^T\boldsymbol{\Gamma}_j k\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{y}_i^T\boldsymbol{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\left\|\boldsymbol{\Gamma}_i\right\|_1$$

Subject to:

$$\sum_{i=1}^{\ell}\boldsymbol{\Gamma}_i = 0, \quad \left\|\boldsymbol{\Gamma}_i\right\|_1 \le C$$

Note: D is non-smooth in its objective and constraint.

The dual problem after introduction of γ_i and γ_i^*:

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\left(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\right)^T\left(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^*\right)k\left(\mathbf{x}_i,\mathbf{x}_j\right)$$

$$+\sum_{i=1}^{\ell}\mathbf{y}_i^T\left(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_i^*\right) - \varepsilon\sum_{i=1}^{\ell}\mathbf{1}^T\left(\boldsymbol{\gamma}_i + \boldsymbol{\gamma}_i^*\right)$$

Subject to:

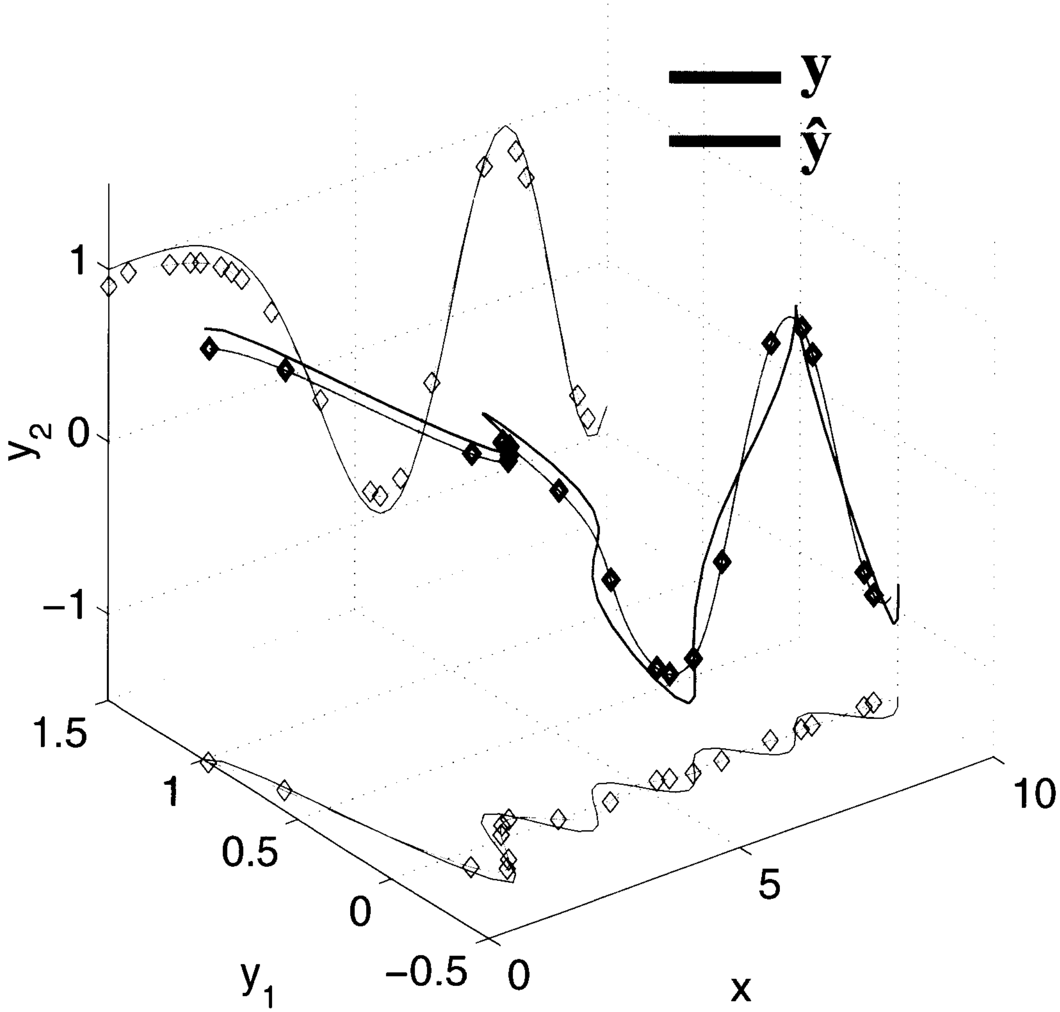
$$\sum_{i=1}^{\ell}\boldsymbol{\Gamma}_i = 0, \quad \mathbf{1}^T\left(\boldsymbol{\gamma}_i + \boldsymbol{\gamma}_i^*\right) \le C, \quad \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i^* \ge \mathbf{0}$$

Note: D is quadratic in its objective and linear in its constraint.

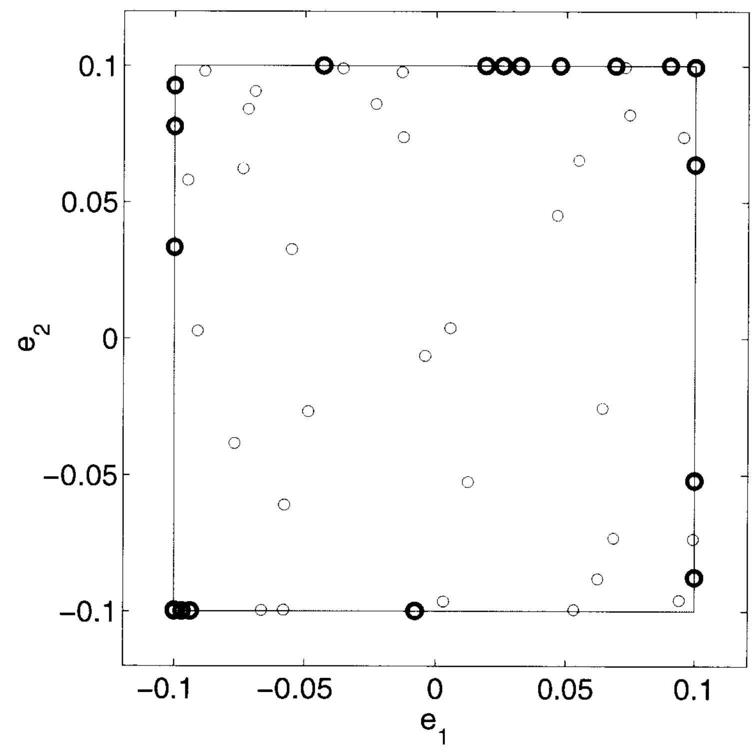
Solve for Γ_i using a standard QP package.

Demonstration (∞-Norm)

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



19 support vectors



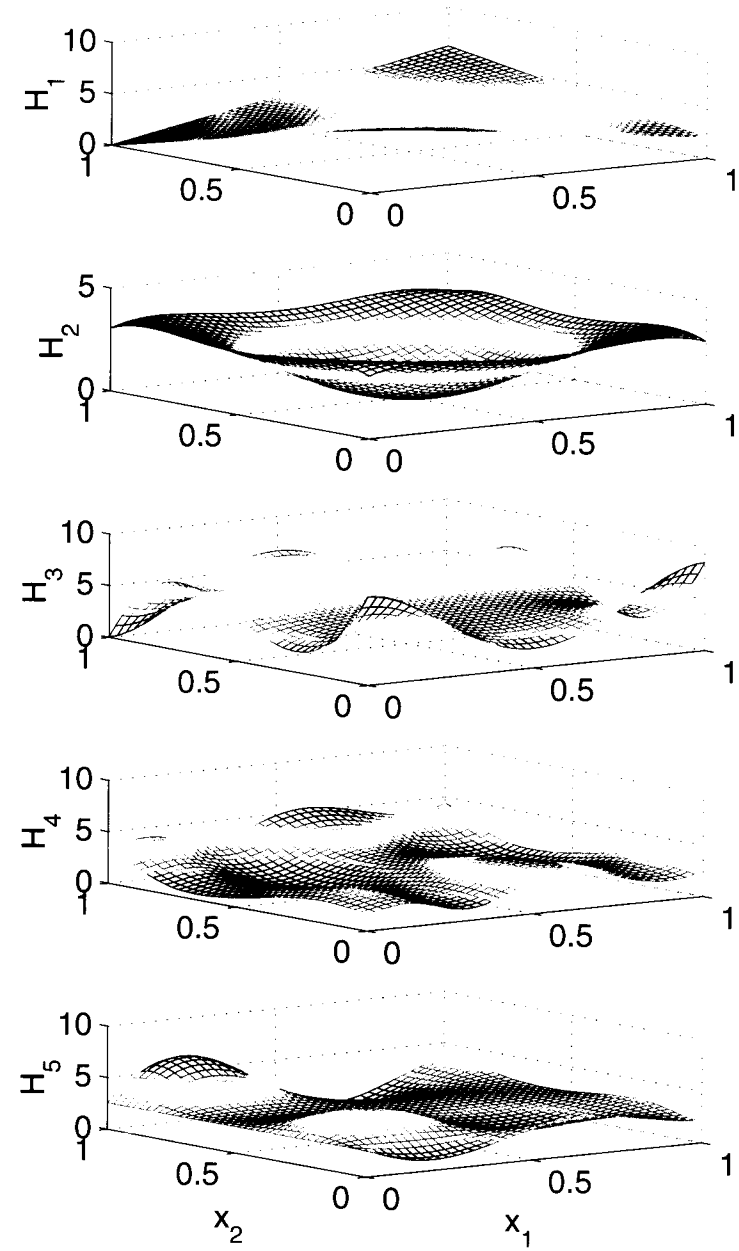
Matlab's `quadprog()` was used to find Γ_i.

Example Hwang

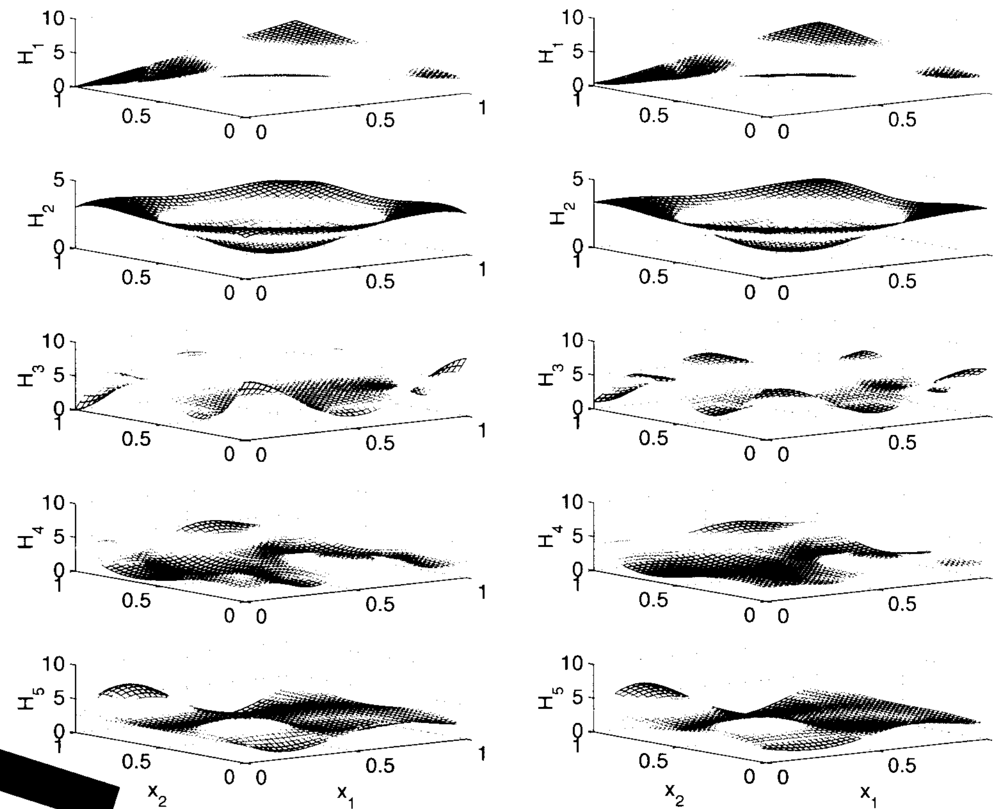
- Hwang data set

$$\mathbf{H}:[0,1]^2 \mapsto \mathbb{R}_+^5$$

- Artificial vector-valued data set
- Of Historical significance
- Input domain is randomly sampled



Hwang



$\mathbf{H}(\mathbf{x})$ $\qquad$ $\hat{\mathbf{H}}(\mathbf{x})$

$$\ell = 500, \quad \varepsilon = 0.5$$

$$C = 100, \quad \gamma = 8$$

$$\left\{\mathbf{x}_i, \mathbf{H}_i\right\}_{i=1}^{\ell}$$

Maximize:

$$D = -\tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\mathbf{\Gamma}_i^T\mathbf{\Gamma}_j k\left(\mathbf{x}_i, \mathbf{x}_j\right) + \sum_{i=1}^{\ell}\mathbf{H}_i^T\mathbf{\Gamma}_i - \varepsilon\sum_{i=1}^{\ell}\left\|\mathbf{\Gamma}_i\right\|_2$$

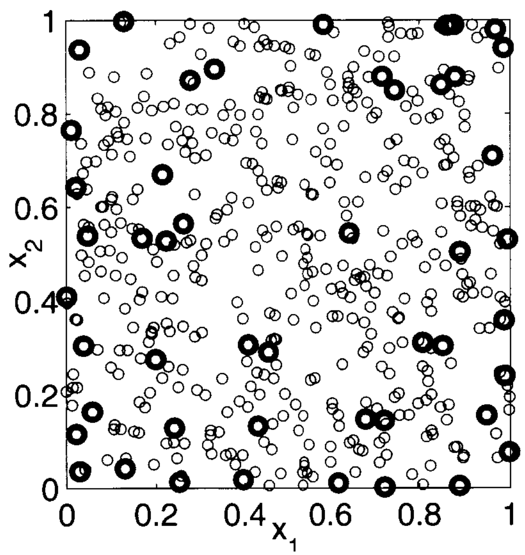
Subject to: $\displaystyle\sum_{i=1}^{\ell}\mathbf{\Gamma}_i = 0, \quad \left\|\mathbf{\Gamma}_i\right\|_2 \le C$

$$\hat{\mathbf{H}}(\mathbf{x}) = \sum_{i \in \mathcal{SV}} \mathbf{\Gamma}_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}$$

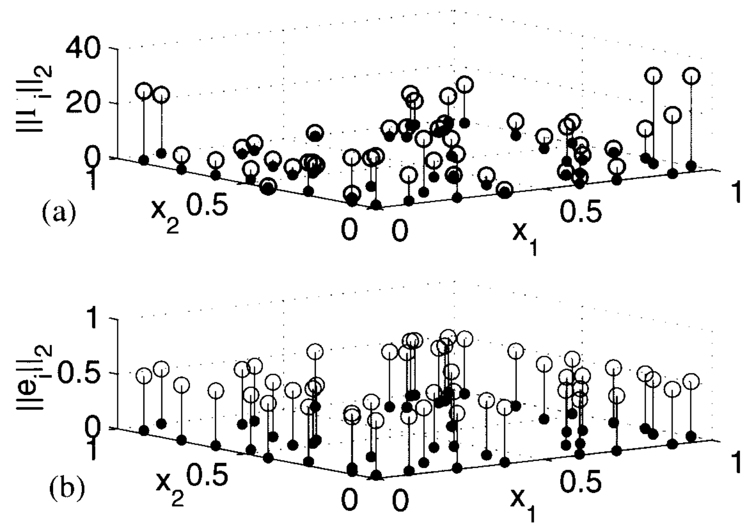
$$k\left(\mathbf{x}_1, \mathbf{x}_2\right) = \exp\left(-\gamma\left\|\mathbf{x}_1 - \mathbf{x}_2\right\|_2^2\right)$$

Hwang KKT

$$\left\|\Gamma_i\right\|_2 \in (0, C) \Rightarrow \left\|\mathbf{e}_i\right\|_2 = \varepsilon$$



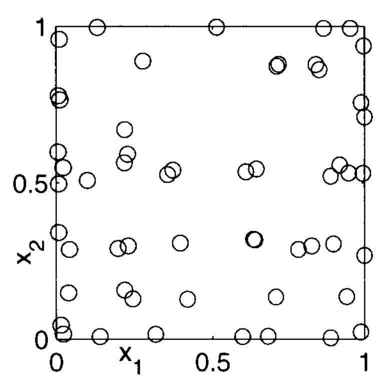
48 Support Vectors

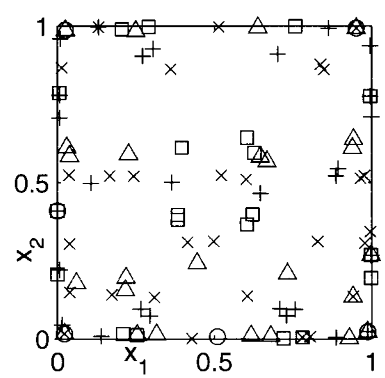


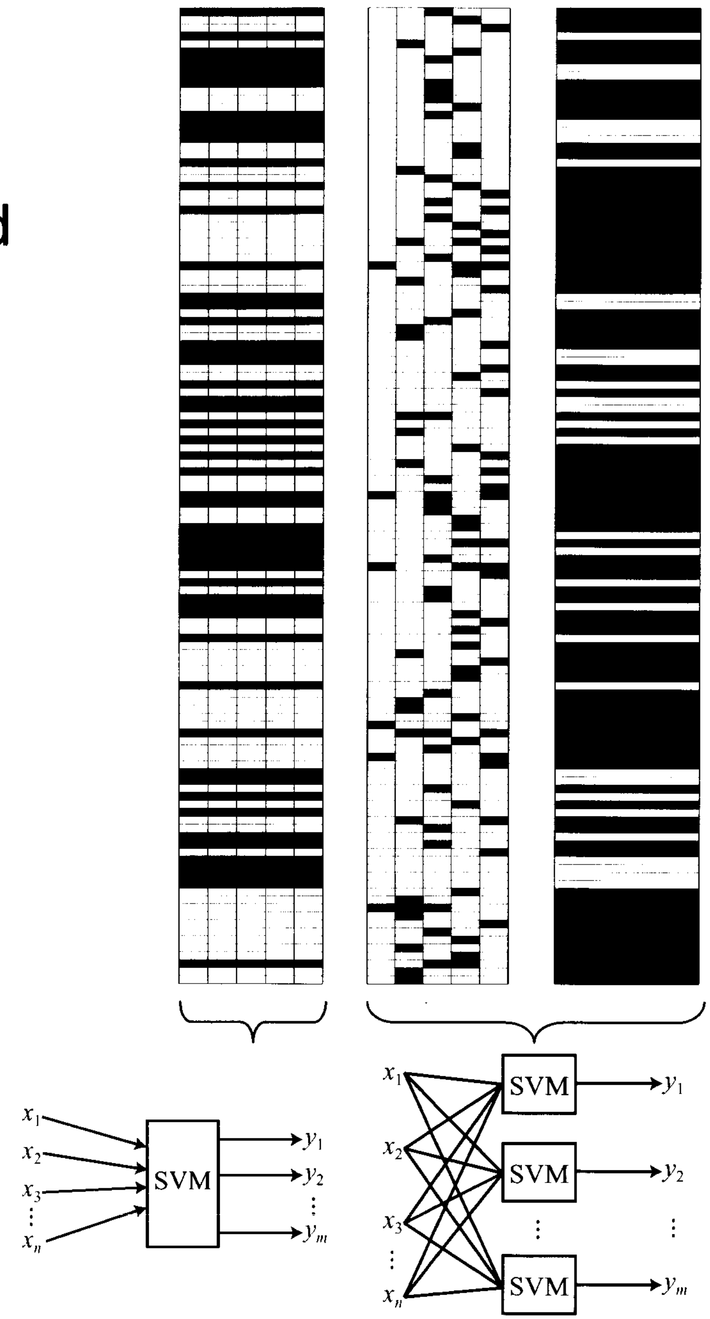
KKT Conditions

Sparsisity

- Compare VV-SVR with aggregated SVR (libsvm)
- Hwang data set with 2,000 points
- Equal Volume
  - VV-SVR $\varepsilon = 0.5$
  - SVR $\varepsilon = 0.34850$
- Support Vectors
  - VV-SVR 55
  - SVR 92







Conclusions

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

- VV-SVR generalizes the scalar-valued case.
  - Estimator form and parameters
  - Loss function
  - Regularization functional

- VV-SVR maintains the sparsity of the scalar-valued case.